2025 Scientific computing environments

Dennis Gannon
Director Cloud Research Strategy

Microsoft Research Connections

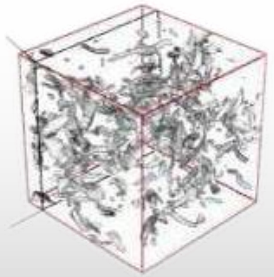
The data explosion is transforming science

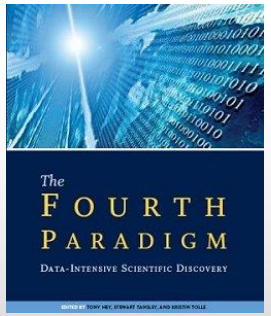
Petabytes
Doubling & Doubling

- Every area of science is now engaged in data-intensive research
- Researchers need
  - Technology to publish and share data in the cloud
  - Data analytics tools to explore massive data collections
  - A sustainable economic model for scientific analysis, collaboration and data curation

The Changing Nature Of Research

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$





Experimental

Thousand years ago

Description of natural phenomena

Theoretical

Last few hundred years

Newton's laws, Maxwell's equations…

Computational

Last few decades

Simulation of complex phenomena

The Fourth Paradigm

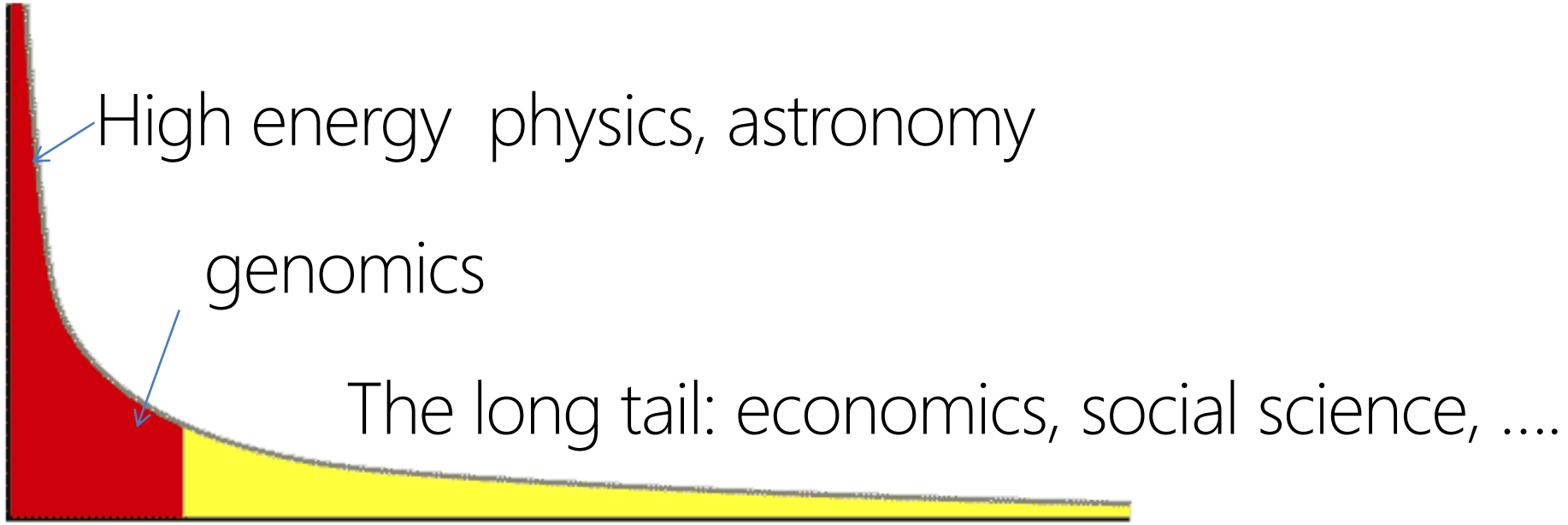
Today and the Future

Unify theory, experiment and simulation with large multidisciplinary Data

Using data exploration and data mining (from instruments, sensors, humans…)

Distributed Communities

The Long Tail of Science



Collectively "long tail" science is generating a lot of data

- Estimated at over 1PB per year and it is growing fast.

Many funding agencies now or soon will requires all data be made public

- US Universities are struggling with this new load
- Data must be preserved
- Data must be sharable, searchable, and analyzable

Bringing Large scale data analytics to more people. Let Scientists Be Scientists…

Most scientists do not want to be system administrators

They don't want to learn to use supercomputers

They want to focus on their science

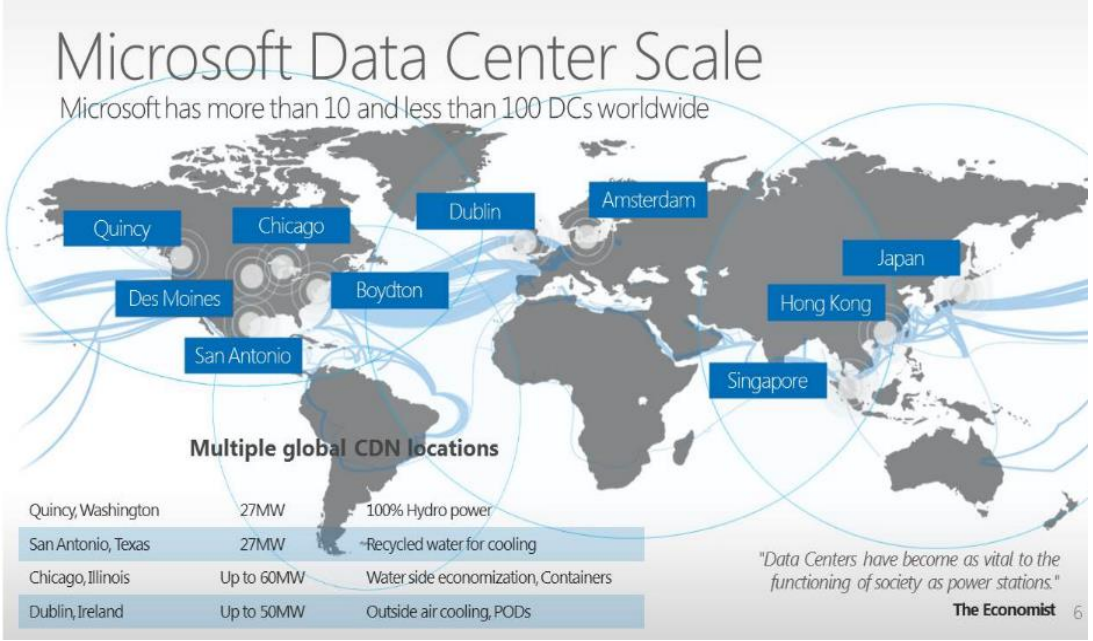
They use standard tools: spreadsheets, statistical packages, desktop visualization

Programming = modifying a few parameters in a trusted scripting language

They want to share experiments with their collaborators

Drivers of Change

- Exascale build-out of commercial cloud infrastructure
  - For 2014-15 expect 10,000,000 new servers and 10 Exabytes of storage in major commercial cloud data centers worldwide.
- A revolution in deep NN machine learning is transforming data analytics.
  - Applications: genome-wide association studies, computer vision and automatic speech translation, medical and environmental image analysis.
  - Evolution of urban science



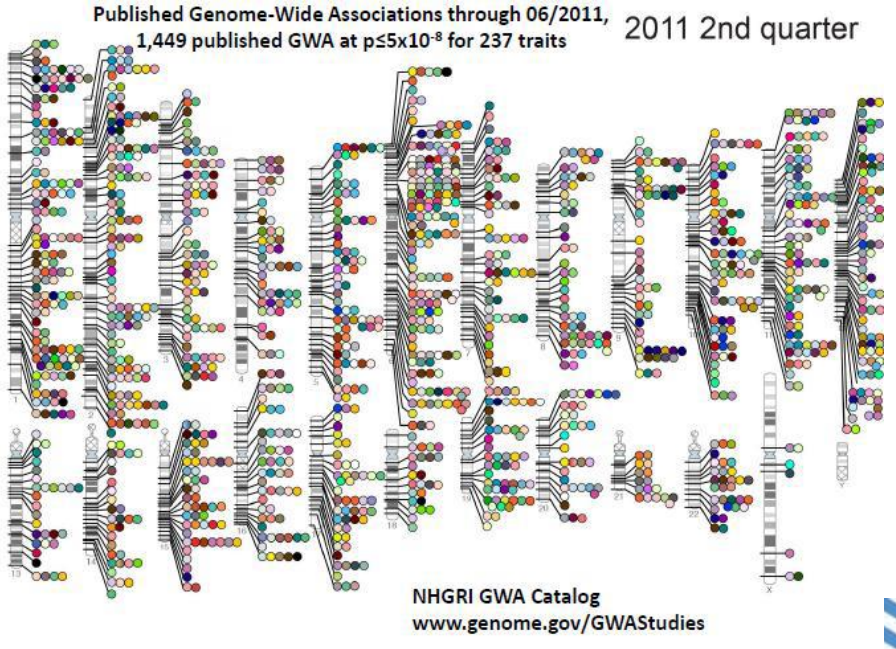
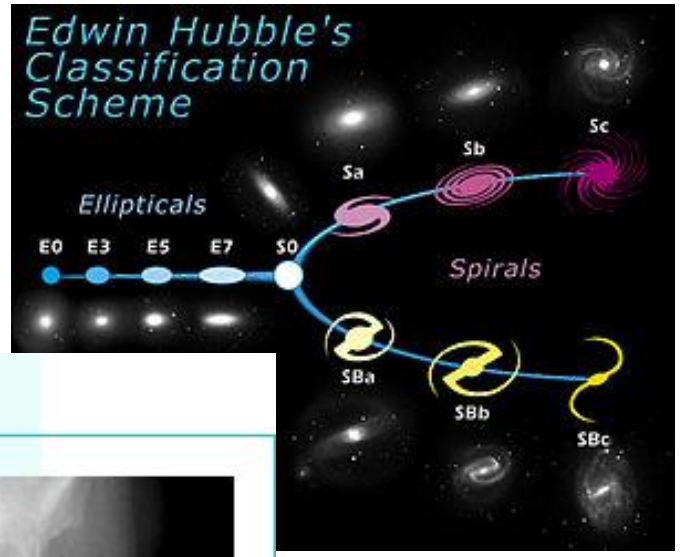
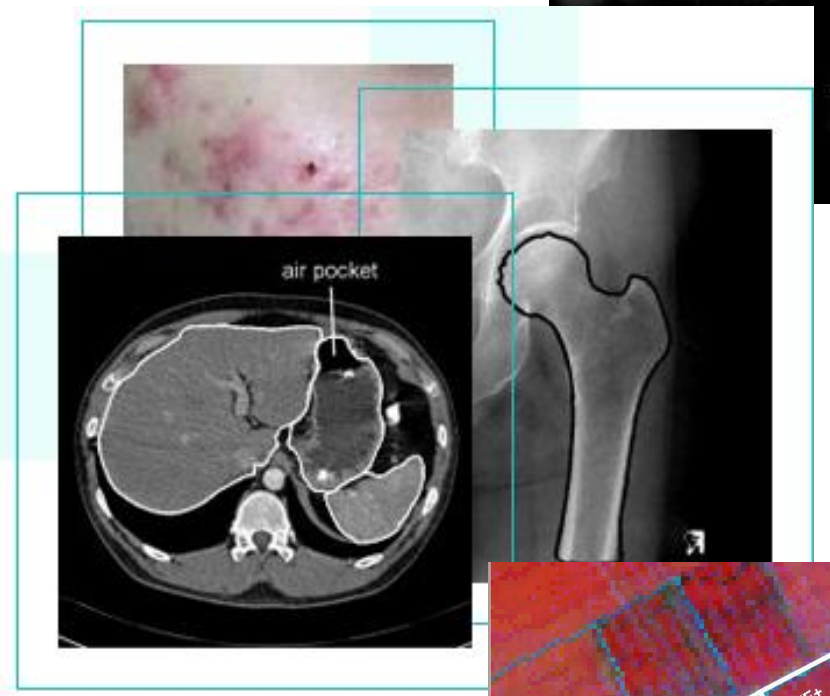


Image Analysis

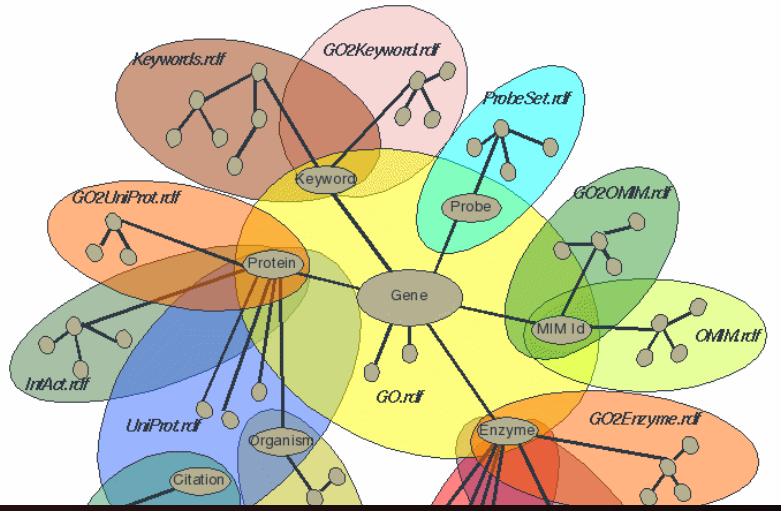
- Astronomy
  - Galaxy classification
  - Spotting planets
- Medical
  - Organ identification
  - Recognizing potential anomalies
- HE Physics
  - Particle collision analysis
- Geosciences
  - Learning to read seismic data
  - Recognizing crop diseases, mineral resources, etc

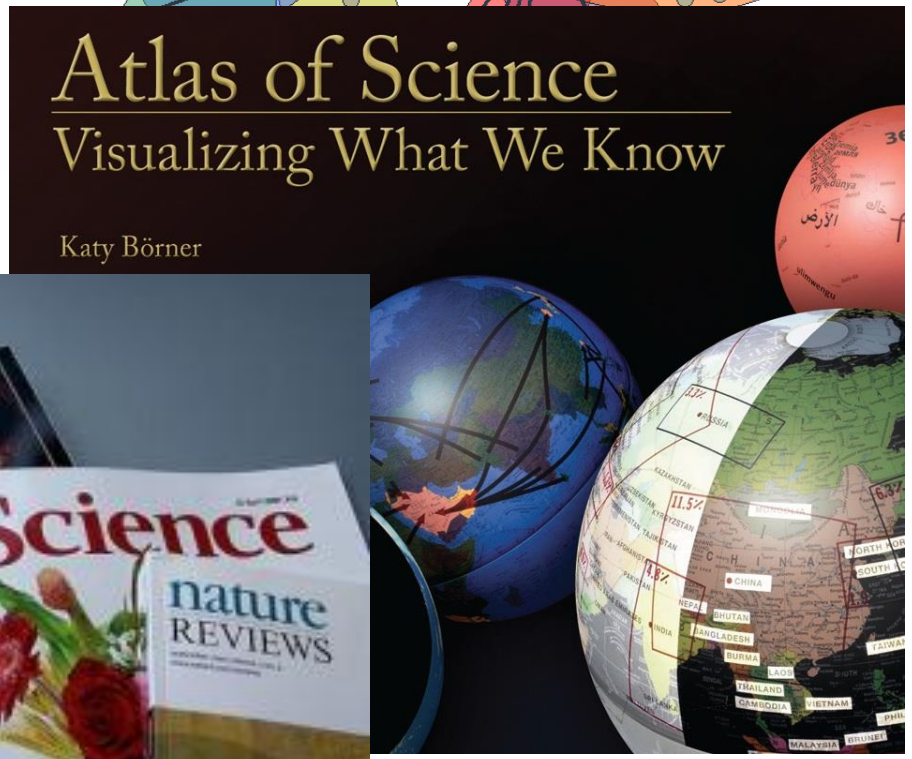




Scientific Document Analysis

- Infer ontologies and classify documents
- Recognize inconsistent scientific results
- Understand/predict the evolution of science
- Reproducible Science
  - Use the digital paper to redo the analysis based on complete workflow and data provenance

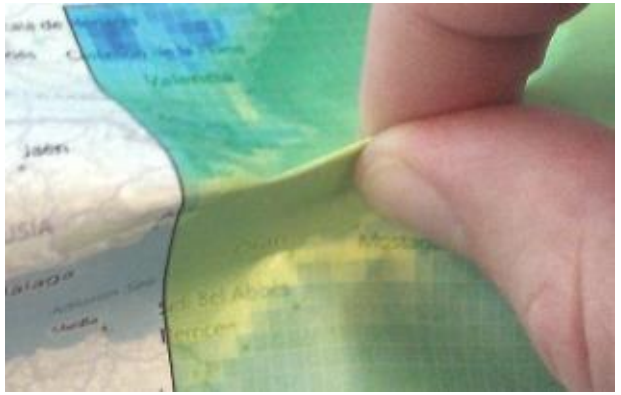




esearch Connections

Cloud apps for science

- New generation of data exploration tools are available now.



FetchClimate

ChronoZoom: An infinite canvas in time

2025 vision of long-tail research

- Scientific data is available to all in the cloud.
- Extrapolate Apple Siri and Wolfram Alpha and Google/Bing …
- Users will interact with it via a computational agent
  - Agent runs on our personal devices and in the cloud
  - We interact with it using voice and gestures
  - The agent understands our research interests and is capable of mining the data universe looking for results of our scientific queries.
- Collaboration will be based on an evolution of social networking tools
  - Domain specific community supporting advanced analytics around high value data archives.
- The environmental impact of research will be fundamental.